



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12Q 1/68, C07H 15/12, C12N 15/00	A1	(11) International Publication Number: WO 92/08117 (43) International Publication Date: 14 May 1992 (14.05.92)
---	----	--

(21) International Application Number: PCT/US91/07308

(22) International Filing Date: 8 October 1991 (08.10.91)

(30) Priority data:
9002259 17 October 1990 (17.10.90) NL

(71) Applicant (for all designated States except US): APPLIED BIOSYSTEMS, INC. [US/US]; 850 Lincoln Centre Drive, Foster City, CA 94404 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): TILANUS, Marcel, G., J. [NL/NL]; Grutteweide 4, NL-6708 BH Wageningen (NL).

(74) Agent: SMITH, Joseph, H.; Applied Biosystems, Inc., 850 Lincoln Centre Drive, Foster City, CA 94404 (US).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.

Published*With international search report.**Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: METHOD OF DETERMINING A GENOTYPE BY COMPARING THE NUCLEOTIDE SEQUENCE OF MEMBERS OF A GENE FAMILY AND KIT THEREFOR

(57) Abstract

A method is provided for determining a genotype by comparing the nucleotide sequence of members of a gene system which flank the polymorphous sections of the loci of interest. This method is characterized in that the nucleotide sequence of the locus or gene of a strongly conserved section is compared in which the polymorphism is localized.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU+	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE*	Germany	MC	Monaco	US	United States of America
DK	Denmark				

+ Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

METHOD OF DETERMINING A GENOTYPE BY COMPARING THE NUCLEOTIDE SEQUENCE OF MEMBERS OF A GENE FAMILY AND KIT THEREFOR

5 The invention relates to a method of determining a genotype by comparing members of a gene family, and also to a kit for tracing genetic variations in a gene family.

A gene system is composed of different loci which are in turn composed of one or more genes; these genes may contain variations, so-called different alleles, for each system. The immunoglobulin supergene family, which also includes, *inter alia*, the T-cell receptor, the immunoglobulin and the HLA (= human leukocyte antigen) gene systems, is characterised by the presence of a large variation (polymorphism). Each of the genes which form part of the gene systems is involved in the immune response in one way or another. Defects in the immune response, which are due to diverse variations in one or more of the gene systems already mentioned, may result in disease.

Genetic variation may thus be associated with diseases or disease symptoms. Identification of loci/ alleles may therefore, for example, be of importance in determining the risk of a disease associated with HLA or in detecting variations in the gene system which result in a deviation, and variations for which the latter is not the case. Thus, the HLA-B27 antigen is/are found in 90% of all patients having rheumatoid spondylitis, and HLA-DR3 or HLA-DR4, or both, is/are found in 80% of patients afflicted with insulin-dependent diabetes. Anyone having HLA-B27 has 180 times as much chance of getting rheumatoid spondylitis as a person without HLA-B27. Table I (taken from Biotechnologie, een nieuwe industriële revolutie (Biotechnology. A New Industrial Revolution), Antebi, E. and Fishlock, D., published by Natuur en Techniek, Maastricht/Brussels (1987) page 91) shows the most important diseases which are known to be associated with the HLA system. These locus and allele associations may render clinical practice a great service. They are sometimes used for diagnosis and sometimes in the treatment of certain diseases. They furthermore open up the possibility of prevention. It is possible to identify those which have the greatest risk of a particular disease. That is very important in those situations in which preventive or early treatment is possible and the development of the disease can be contained.

Table I
The most important diseases associated with the HLA system

5	HLAs	Pathology	Relative risk*
	A1	Hodgkin's disease	1.4
	A3	Idiopathic haemochromatosis	8.2
	B5	Behçet's disease	6.3
10	B14, B47	Congenital adrenal hyperplasia	15.4
	B27	Rheumatoid spondylitis	87.4
	B35	Subacute thyroiditis	13.7
	Cw6	Psoriasis	13.3
	DR2	Multiple sclerosis	4.1
15	DR3	Systemic lupus erythematosus	5.8
	DR3	Addison's disease	6.3
	DR3	Basedow's disease	3.7
	DR3	Myasthenia gravis	2.5
	DR3	Extramembranous glomerulopathy	12.0
20	DR3 and/or DR4	Insulin-dependent diabetes	6.4
	DR3 and/or DR7	Coeliac disease	4.8
	DR4	Rheumatoid arthritis	4.0
	DR4	Pemphigus	14.4
	DR5	Biermer's anaemia	5.4
25	DR5	Hashimoto's disease	5.6
	DR5, DRw8	Juvenile arthritis	5.2

* The relative risk indicates the factor by which the risk of getting a disease is multiplied for people having the HLAs concerned.
30 Individuals having group HLA-Cw6 have, for example, 13.3 times as much chance of getting psoriasis as people not having said antigen.

Genetic variation in the HLA system is furthermore of importance in view of the fact that acceptance in organ and tissue transplants is 35 associated with the presence of the same HLA alleles in the white blood cells, while rejection is related to the presence of various HLA alleles in the white blood cells.

In man the HLA gene family covers a region of 1,500,000 base pairs. This region codes for many genes. Thus, this region contains, in addition 40 on to the genes of HLA class I and HLA class II, inter alia, genes of

the complement system, 21-hydroxylase genes and genes which code for tumour necrosis factor.

The HLA class I region contains 3 "major" loci: HLA-A, HLA-B and HLA-C. The molecules for which said loci code are expressed in all 5 nucleated cells. The molecules of class I are composed of a polymorphous α -chain having a molecular weight (MW) of 44,000 D and a constant chain, called β_2 -microglobulin, having a MW of 12,000 D.

The HLA class II region contains the loci HLA-DR, HLA-DQ and HLA-DP. In addition to said loci, a number of genes are also described 10 which do not belong to one of the abovementioned loci and are for the time being considered to be separate loci (for example, HLA-E; HLA-F; HLA-DN; HLA-DZ). The molecules of class II are composed of an α -chain (MW 34,000 D) and a β -chain (MW 29,000 D). The α -chain of HLA-DR is constant, i.e. no variants have yet been found. HLA-DR contains 15 3 β -genes, of which two genes code for a product and one gene is a pseudogene which is therefore not expressed. Each of the β -chains can form dimeric molecules with the α -chain. The HLA-DQ locus contains four (two α - and two β -) genes, of which two are expressed, namely DQ α and DQ β (termed DQA and DQB on the recommendation of the latest nomenclature committee). The molecule of the other genes (DXA and DXB) has not been 20 found, although the structure of said genes indicates no characteristic pseudogenes. The HLA-DP locus contains four genes (two α and two β), of which two are expressed. Polymorphism has not yet been demonstrated for DPA, and only a few alleles of DPB have been described. The other two DP 25 genes are pseudogenes.

A method of detecting specific nucleotide variations and genetic polymorphisms which are present in nucleic acids, and also a kit for carrying out said method are described in EP-A-237,362. Said method comprises amplifying nucleic acid sequences in which the presence of 30 nucleotide variations, mutations and polymorphisms is assumed, followed by the detection thereof with allele-specific oligonucleotide sequences using a dot blot. In this process, allele-specific oligonucleotide sequences are used (see page 22 of the said patent application) which comprise at least the DNA sequence with the variation to be detected and 35 have to be specific for the nucleotide variation to be detected. The kit therefore comprises many different primers which code for diverse known variations in the DNA sequence, to be investigated, of the gene system.

Immobilized sequence-specific oligonucleotide probes for the genetic analysis of amplified DNA are disclosed by the International Patent 40 Application WO 89/11548 and by R. K. Saiki et al., Proc. Natl. Acad.

Sci. USA, 86, 6230-6234 (1989).

A method of determining genotypes as regards the alleles in the HLA-DP loci by hybridising nucleic acid samples with a series of probes which are specific for various segments is disclosed by the International Patent Application WO 89/11547.

A method of distinguishing nucleotide sequences by hybridising amplified DNA with linkable probes is disclosed by the International Patent Application WO 90/01563.

These known methods have the following disadvantages:

10 a) It is only possible to trace known variations in the DNA sequence of a particular gene system, as a result of which any hitherto unknown variations in the gene system to be investigated cannot be discovered by said method.

15 b) It is necessary to include various oligonucleotides for every known variation in the kit and to use them in the method, with the result that, in complex gene systems which comprise many possible alleles (such as, for example, the HLA system), a different oligonucleotide has to be used for every possible allele and a different hybridisation has to be carried out, which is cumbersome, time-consuming and,

20 expensive.

HLA TYPING

Class I serology

The HLA class I molecules are determined with the aid of antisera which are specific for individual alleles. In most cases, alloantisera are used because monoclonal antibodies against all the allele forms are not available. It is improbable that they will become available in the short term because the variation which characterises the polymorphism of the various alleles is localised in a small region of the class I gene.

30 Class I protein

Protein-chemistry determination of the class I alleles is possible by means of immune precipitations with monoclonal antisera against constant sections of the molecules, followed by isoelectric focusing (IEF).

35 Class I DNA

Typing is possible to a limited extent by means of determinations of restriction fragment length polymorphism (RFLP) in which use is made of class I cDNA probes. Further refinement has yielded probes which are locus-specific, as a result of which clusters of alleles and some 40 separate alleles can be detected. Typing of class I alleles by means of

allele-specific oligohybridisations (ASO) is not considered possible as a result of the large number of different alleles and, consequently, the large number of oligos which is necessary. Routine typing at the DNA level is difficult to achieve by this method.

5 Class II serology is more difficult to carry out in practice, firstly because of the lack of well-defined antisera and secondly because of the elaborateness of the identification technique, viz. two-colour fluorescence.

10 Class II proteins can be characterised by means of IEF as a result of the availability of monoclonal antibodies, but the diversity of alleles in terms of variation and number makes the standard typing difficult, especially in heterozygous individuals.

Class II DNA

15 RFLP analyses can be used to identify DRA, DQA, DQB and DPB alleles. In this connection, some alleles are placed in a "supertypic group" because they cannot be distinguished individually, but stand out as a group of alleles having a specific RFLP pattern from the other alleles.

20 The technique of Southern blot analyses resulting in RFLP patterns is readily usable for typing class II. Viewed technically, this approach is not suitable for routine class II typing because of its elaborateness.

25 The development of ASO typing with synthetic oligonucleotides on amplified DNA by means of the polymerase chain reaction (PCR) has now made routine typing of DQA possible. Typing will have to be developed still further for DRB, DQB and DPB. A disadvantage of this method is that a large number of allele-specific oligos has to be used as the number of alleles of the locus concerned increases.

Workshop

30 Every year the results of the HLA typing centres distributed around the world are examined at a workshop, sera being selected and defined as reference sera. The improvements in typing by means of protein-chemistry determinations, DNA/RFLP and DNA-ASO analysis is facilitating the standardisation of alleles determined with antisera.

35 As regards protein typing, the workshop determines which monoclonal antibodies have to be used and how the data have to be interpreted. For DNA-RFLP typing, the enzyme-probe combination is determined, in addition to a standardisation of DNA isolations and hybridisations.

40 DNA-ASO typing is offered as a package by Perkin-Elmer/Cetus Corporation (patented and marketed as a kit) and can now be used for DQA

typing.

Attempts are being made to find a combination of the various methods which can be carried out uniformly by every typing centre throughout the world. In this connection, the reproducibility of the 5 determinations in various typing centres, the uniformity of reagents (sera, probes, oligos) and the possibility of carrying out quality controls of the typing play an important role.

In the abovementioned cases, the approach to typing is based on the characterisation of the polymorphous sections of the molecules or genes 10 of the loci.

It has now been found that use can advantageously be made of conserved sequences which flank the polymorphous system, to determine genetic variations.

The method according to the invention for determining a genotype by 15 comparing the nucleotide sequence of members of a gene system is therefore characterised in that the nucleotide sequence of a locus or gene of a strongly conserved section is compared in which the polymorphism is localised the genetic material.

An important advantage of the method according to the invention is 20 that deviations in a gene system, and especially deviations which are expressed, can be traced more easily and with greater certainty by means of it. More particularly, the advantages of the present method are:

- New mutations can be detected directly and can then be investigated in more detail for functionality. This is in contrast to serological 25 typing, in which at most a cross reaction with other sera but more often no reaction, is found. ASO typing characterises them as negative for that allele.

The method is independent of known polymorphisms compared with other determinations.

30 - Owing to the simplicity of the system (high degree of reproducibility throughout the world), standardisation of typing will be simplified.

- The method is independent of geographically determined internal variations in blood cells, serum and expression of the alleles.

35 - At the same time the alleles of different loci can be determined in one test, as a result of which the association with diseases becomes known in a broader context (more loci or gene systems).

- The nomenclature for the gene systems can be localised in a simple manner by definitions of polymorphous parts of the loci/genes by using 40 the invention.

The strongly conserved section of the genetic material whose nucleotide sequence is compared, according to the invention, has a larger homology than the polymorphous (less conserved) sections whose sequence is compared in the known methods. The homology at base level, i.e. the 5 percentage of nucleotides which is characteristic of a locus or of the allele forms of the genes is at least approximately 50% and preferably at least 58%.

The method according to the invention is especially usable for determining haplotypes of a gene system.

10 In particular, the method can be used to compare loci of a gene system and alleles of a gene system, such as the gene system of the histocompatibility system. At the same time, it may be a matter of a routine determination of alleles in which a specific typing/identification of the presence of an allele is required (for example, in bone 15 marrow transplants or blood transfusions).

A gene system which is also advantageously suitable for investigation with the method according to the invention is a gene system of one of the members of the immunoglobulin supergene family, in particular the genes of an immunoglobulin (Ig) or the genes of the T-cell receptor. 20 (TCR).

The method according to the invention is also suitable for research purposes in which a simpler determination of polymorphous sequences other than those which are characteristic for typing is possible in a gene as a result of selection of other conserved flanking sequences.

25 The determination of the genotype with the method according to the invention relates, in particular, to tissue typing or cell typing. This can be done, for example, in order to determine the degree of compatibility of transplants. If no, or very few, deviations are found in the polymorphous region between donor and acceptor, a transplant will have a 30 high chance of taking because no rejection takes place on the basis of HLAs foreign to the body. It is also possible to carry out tissue typing or cell typing in order to determine the risk of an HLA-associated disease for an animal. A particular aspect in this connection is the uniform 35 performability and introduction of typing possibilities for, for example, HLA class I alleles which could hitherto be determined only by serological techniques.

Moreover, a gene system of a bacterium, of a virus or of lipoproteins can be used as the gene system whose genetic variation is investigated.

40 The investigation of the possible variation in the nucleotide

sequence of a strongly conserved region with the aid of the method according to the invention can readily be carried out if the strongly conserved section to be investigated is bounded on at least one side, and preferably on both sides, by a completely conserved sequence, i.e. a 5 sequence which is completely identical in all individuals. Said completely conserved sequences have, in particular, a length of at least 5 nucleotides and, particularly preferably, a length of at least 10 nucleotides.

The strongly conserved nucleotide sequences to be investigated can 10 be amplified using the completely conserved sequences. A method which is used with advantage for amplifying the strongly conserved nucleotide sequence is a PCR (polymerase chain reaction) (Saiki R.K., Gelfand D.H., Stoffel S., Scharf S.J., Higuchi R., Horn G.T., Mullis K.B. and Erlich H.A. (1988). Primer directed enzymatic amplification of DNA with a 15 thermostable DNA polymerase, *Science* 239, 487-491). Other methods of amplifying the sequence to be investigated are also usable in the method according to the invention.

The sequence amplified in this way can then be subjected to direct 20 sequence analysis, to denaturing gradient gel electrophoresis (DGGE) or to temperature-gradient gel electrophoresis (TGGE). The extent to which the denaturing gel system works depends on the discriminating capacity of the gel in relation to the number of different alleles. In other words, the gel system will work well if there is only a limited number 25 of alleles of a locus, such as the lipoprotein alleles, HLA-DQA alleles, HLA-DPB alleles and possibly HLA-DQB alleles, and virus types, and can be used to determine expression of different loci in different cells or tissues.

The amplified DNA or RNA can also be analysed (= typed) by direct 30 sequence determination methods. For routine typing determinations, both analyses may be carried out informatively, reproducibly and in a controlled manner. For gene systems containing several alleles, direct sequence analysis will facilitate the interpretation and therefore the reproducibility by the various centres. The direct sequence analysis method also has advantages for research purposes.

35 Determination of modifications has hitherto been carried out at DNA level by means of Southern blotting. The different families of TCR and Ig genes can be detected by means of specific amplification and sequencing or DGGE, while the various members of the family can be distinguished by a set of family-specific primers. Another gene system which 40 is directly suitable for this approach is the determination of diverse

viruses, for example HPV1-16.

Selection of primers

Many alleles of a large number of genes have been sequenced for research purposes. These sequences are stored in the EMBL databank (for Europe) and in the Gene bank (USA), both of which are accessible to subscribers. On the basis of these sequences, the allele-specific oligos (ASO), for example, have been selected from the polymorphous regions and these have then been used to identify different alleles. Most gene systems contain regions of sequences which are not subject to, hardly subject to or very strongly subject to mutations. For the HLA class II gene system, this has been investigated in detail in order to localise the polymorphous regions. References: Gustafsson K., Emmoth E., Larhammar D., Bohme J., Hyldig-Nielsen J.J., Peterson P.A. and Rask L. (1984), Mutations and selection in the generation of class II histocompatibility antigen polymorphism. *Embo J.* 3, 1655-1661; Gustafsson K., Wiman K., Larhammar D., Rask L. and Peterson P.A. (1984), Signal sequences distinguish class II histocompatibility antigen beta chains of different loci. *Scand. J. Immunol.* 19, 91-97.

The presence of usable sequences for the method according to the invention can be detected by more detailed analysis of HLA class I and II sequences which were available from the sequence databanks (EMBL and Genebank), supplemented by self-determined sequences. Primers for class I and other gene systems have been selected from the sequence databank and can be used. Primers based on conserved sequences for the HLA class II loci DRB, DQA, DQB and DPB can be selected on the basis of the sequences already known. Some usable primers are reported in one of the examples below. It is within the scope of the person skilled in the art to select and to test the completely conserved sequences (primers) suitable for the typing he plans.

Applications

To determine lipoprotein polymorphism, the ASO can be radioactively labelled and used as a probe for typing the various alleles which are present in lipoprotein E. Three frequently occurring alleles can also be determined by means of restriction enzyme analysis because, in the case of these three alleles in the polymorphous section, the differences are localised at restriction enzyme sites. All five different alleles can be determined directly by a denaturing gel electrophoresis (Sheffield et al., Identifying DNA polymorphisms by denaturing gradient gel electro-

phoresis, in PCR Protocols: A Guide to Methods and Applications, Academic Press Inc., 1990, pages 206-218). There are indications that the different alleles can be detected by direct sequencing since the occurrence of the different alleles is based on differences in nucleotide sequences (McBride et al., *Cin. Chem.* 35, 2196-2201 (1989)).

In the case of HLA typing, "constant primers" were used to amplify DRB genes of a family in which an abnormal HLA type occurred. After amplification of the DRB genes, cloning in vectors and subsequent sequence analysis, it was found that the "abnormal" HLA type is in fact 10 a new haplotype which cannot be detected by known techniques. In this analysis it was found that the identification of the various alleles can be carried out simply by direct sequencing and localisation of the thymidine nucleotides present in the allele (T-tracking). Automatic sequence determinations have already been carried out for the sequencing 15 of HLA-DQA (McBride et al., 1989, see above). The method followed by them can be used directly with slight modifications for automatic sequence systems to type the alleles of a locus.

After the primers have been selected, "kits" can be assembled for the various gene systems, which kits contain, *inter alia*, said primers, 20 as a result of which a standardised method is obtained for typing the various alleles. In the first instance, the ApoE and HLA-DRB, DQA DQB and DPB alleles are investigated for general applicability with denaturing gels and by means of sequence analysis.

The invention therefore also relates to a kit for tracing genetic 25 variations in a gene family which contains one or two completely conserved sequences having a length of at least 5 nucleotides as primer(s) and possibly other means for amplifying and analysing a strongly conserved section of the genetic material to be investigated, which section is flanked by the completely conserved nucleotide 30 sequence(s).

EXAMPLE 1

Primers based on conserved sequences have been selected for the HLA class II loci DRB, DQA, DQB and DPB and tested for usability. Shown 35 below are the sequences of some universal primers which have been selected and tested for an amplification reaction (PCR). Two primers are preferably used for each reaction.

DRB5: 5' CC AGC ACG TTI C 3'

DRB3: 5' TT GTR TCT GCA 3' (R = G/A)

DQA5: 5' CCA TGA ATT TGA TGG AGA 3'

5 DQA3: 5' ATC GTT TAA TCA 3'

DQA3.1: 5' TGT TCA AGT TRT GTT TT 3' (R = G/A)

DQA3.2: 5' ACA GCS ATG TTT STC AGT GCA 3' (S = G/C)

N.B. DQA5 is used in combination with DQA3, DQA3.1 or DQA3.2.

10 DXA3.2: 5' ACA GCS ATA TTT STC AGT GCA 3' (S = G/C)

DXA3: 5' TCT GCS GGT CAA AAC T 3'

N.B. DXA3 and DXA3.2 are used in combination with DQA5.

DQB5: 5' TGT GCT ACT TCA CCA A 3'

15 DQB3: 5' GTC GTT GTG TCT GCA 3'

DPB5: 5' GCA GGA ATG CTA CGC GTT TA 3'

DPB3: 5' CCA GCT CGT AGT TGT GTC TGC 3'

DPB3SP (alternative for DPB3):

20 5' TTG TAA AAC GAC GGC CAG TCC AGC TCG TAG TTG TGT CTG C 3'

APOE.F4:5' ACA GAA TTC GCC CCG GCC TGG TAC AC 3'

APOE.F6:5' TAA GCT TGG CAC GCC TGT CCA AGG A 3'

25 The combinations DRB5/DRB3, DQA5/DQA3.2 and DPB5/DPB3SP have been successfully used for sequences.

EXAMPLE 2

In the family G (two parents, four children) cross reaction was found in the determination for HLA class II alleles having different 30 typing sera by means of the conventional HLA typing method. It was not possible to establish any typing by this means. More detailed analysis by means of restriction fragment length polymorphism (RFLP) and ASO typing resulted in the hypothesis that it was not two alleles of HLA-DRB which occurred in this family in different members of the family, but 35 three, and this had implications for a disease of one of the members of the family to be treated (Tilanus M.G.J., Van Eggermond M.C.J.A., Fei H., Scheuder G.M.Th. and Giphart M.J. (1989), Family with an apparent HLA-DR triplet: Evidence for exchange of functional HLA-DR beta genes between different haplotypes, Exp. Clin. Immunogenetics 6, 162-168). The 40 DNA of this family was studied in more detail in order to be able to

establish the genetic basis for this phenomenon. One of the techniques which was used in doing this consisted of sequence analysis of the various alleles of HLA class II loci.

In this family, three DRB alleles have been identified in the DNA 5 of one person, viz. DR1, DR2 and DRw6, on the basis of positive identification with allele-specific oligos (triplet). The allele-specific oligos detect the corresponding sequences for the regions which also differ at protein level. Figure 1 shows the nucleotide sequence of DR1 with the derived amino acid sequence in three-letter and one-letter 10 code. The underlined regions indicate the location of allele-specific oligos. The nucleotide sequence and the associated amino acid sequence of the allele-specific oligonucleotides of the other alleles present in this family are shown in Figures 2a and 2b. Figure 2a shows the nucleotide variation in a first polymorphous region and Figure 2b that in a 15 second polymorphous region, corresponding to the underlined regions in Figure 1. For a more detailed analysis of the DRB genes, the universal primers DRB5S and DRB3E described were used. In a control study, the use of the universal primers for identification of the DRB alleles DR1, DR2 and DRw6 was tested, but no deviations were found in these sequences. The results of the sequence analysis of control cell lines and DNA 20 of family G are shown in Figures 3a and 3b, respectively, in the form of the amino acid sequence using the one-letter code.

It is evident from this that there is a recombined haplotype in this family, DR1 and DR2 alleles being involved. The DRw6,w19 and 25 DRw6,52^c alleles are identical to the control sequences. It appears that the alleles which are detected by the DR1-specific oligos and by the DR2-specific oligo occur next to each other (Figure 4). There is therefore no question of three alleles being present here but of a DRB haplotype not previously identified. In another section of the gene, a specific sequence has been identified for this DR1 + 2 haplotype. 30

Of other loci (DQB, DQA and DPB) in this family, it has now been established by using universal primers that there are also only two (unmutated) alleles. In addition, said haplotype can be characterised by the presence of conserved specific sequences.

CLAIMS

1. Method of determining a genotype by comparing the nucleotide sequence of members of a gene system, characterised in that the nucleotide sequence of a strongly conserved section of the genetic material is compared.
2. Method according to Claim 1, characterised in that the strongly conserved section of the genetic material has a homology of at least 58%.
3. Method according to Claim 1 or 2, characterised in that haplotypes of a gene system are determined.
4. Method according to Claim 3, characterised in that loci of a gene system are compared.
5. Method according to Claim 4, characterised in that alleles of a gene system are compared.
- 15 6. Method according to any of Claims 3-5, characterised in that a gene system of the histocompatibility system is used as gene system.
7. Method according to Claim 1 or 2, characterised in that a gene system of one of the members of the immunoglobulin supergene family is used as gene system.
- 20 8. Method according to Claim 7, characterised in that the gene system of an immunoglobulin is used as gene system.
9. Method according to Claim 7, characterised in that the gene system of the T-cell receptor is used as gene system.
10. Method according to Claim 1 or 2, characterised in that tissue 25 typing or cell typing is carried out.
11. Method according to Claim 10, characterised in that the degree of compatibility of transplants or cells is determined.
12. Method according to Claim 10, characterised in that the risk is determined for a person or animal of a disease associated with HLA 30 (human leucocyte antigen) or MHC (major histocompatibility complex).
13. Method according to Claim 1 or 2, characterised in that a gene system of a bacterium or of a virus is used as gene system.
14. Method according to any of the preceding claims, characterised in that the strongly conserved section is flanked by completely conserved 35 sequences of at least 5 nucleotides.
15. Method according to Claim 14, characterised in that the completely conserved sequences contain at least 10 nucleotides.
16. Method according to Claim 14 or 15, characterised in that the strongly conserved nucleotide sequence is amplified making use of the 40 completely conserved sequences.

17. Method according to Claim 16, characterised in that the strongly conserved nucleotide sequence is amplified using a PCR (polymerase chain reaction).
18. Method according to Claim 16 or 17, characterised in that the 5 amplified sequence is subjected to direct sequence analysis.
19. Method according to Claim 16 or 17, characterised in that the amplified sequence is subjected to denaturing gradient gel electrophoresis (DGGE).
20. Method according to Claim 16 or 17, characterised in that the 10 amplified sequence is subjected to temperature-gradient gel electrophoresis (TGGE).
21. Kit for tracing genetic variations in a gene family which contains one or two completely conserved nucleotide sequences having a length of at least 5 nucleotides as primer(s) and possibly other means for amplifying and analysing a strongly conserved section of the genetic material 15 to be investigated, which section is flanked by the completely conserved nucleotide sequence(s).

Fig. 1

DR1

CGTTTCTTGTGGCAGCTTAAGTTGAATGTCATTCCTCAATGGGACGGAGCGGGTGCAG	32	62
ArgPheLeuTrpGlnLeuLysPheGluCysHisPhePheAsnGlyThrGluArgValArg		
R F L W Q L K F E C H F F N G T E R V R		
TTGCTGGAAAGATGCATCTATAACCAAGAGGGAGTCCGTGCCCTCGACAGCGACGTGGGG	92	122
LeuLeuGluArgCysIleTyrAsnGlnGluGluSerValArgPheAspSerAspValGly		
L L E R C I Y N Q E E S V R F D S D V G		
GAGTACCGGGCGGTGACGGACCTGGGGCGGCTGATGCCGACTGGAAACGGCAGAAG	152	182
GluTyrArgAlaValThrGluLeuGlyArgProAspAlaGluTyrTrpAsnSerGlnLys		
E Y R A V T E L G R P D A E Y W N S Q K		
GACCTCCTGGAGCAGAGGCAGGGCCGGTGGACACCTACTGCAGACACAAC	212	242
AspLeuLeuGluGlnArgArgAlaAlaValAspThrTyrCysArgHisAsnTyrGlyVal		
D L L E Q R R A A V D T Y C R H N Y G V		
GGTGAGAGCTTACAGTGCAGCGCGAA		
GlyGluSerPheThrValGlnArgArg		
G E S F T V Q R R		

Fig. 2a

DR1

CGT TTC TTG TGG CAG CTT AAG TTT GAA TGT C
R F L W Q L K F E C

DR2-1

CGT TTC CTG TGG CAG CCT AAG AGG GAG TGT C
R F L W Q P K R E C

DR2-5

CGT TTC TTG CAG CAG GAT AAG TAT GAG TGT C
R F L Q Q D K Y E C

DRw6.w19

CGT TTC TTG GAG TAC TCT ACG TCT GAG TGT C
R F L E Y S T S E C

DRw6.52C

CGT TTC TTG GAG CTG CTT AAG TCT GAG TGT C
R F L E L L K S E C

Fig. 2b

DR1

TTG CTG GAA AGA TGC ATC TAT AAC CAA GAG
L L E R C I Y N Q E

DR2*1

TTC CTG GAC AGA TAC TTC TAT AAC CAG GAG
F L D R Y F Y N Q E

DR2*5

TTC CTG CAC AGA GAC ATC TAT AAC CAA GAG
F L H R D I Y N Q E

DRw6.w19

TTC CTG GAC AGA TAC TTC CAT AAC CAG GAG
F L D R Y F H N Q E

DRw6.52C

TTC CTG GAC AGA TAC TTC CAT AAC CAG GAG
F L D R Y F H N Q E

Fig. 3a

Control DNA

DR1	RFLWQLKFECHFFNGTERVRLLERCIYNQEEESVRFD	
DR2*1	----P-R-----F-D-YF-----	
DR2*5	---Q-D-Y-----F-H-D-----DL---	
DRw6.w19	---EYSTS-----F-D-YFH----N---	
DRw6.52 ^c	---EL--S-----F-D-YFH---F---	

Fig. 3b

DNA family G

DR1	-----	
DR1-like	-----F-H-D-----DL---	
DR2-like	---E-A-C---I-----F-H-D-----DL---	
DRw6.w19	---EYSTS-----F-D-YFH----N---	
DRw6.52 ^c	---EL--S-----F-D-YFH---F---	

- = same amino acid as in DR1 at that position.

* = regions for ASO typing

Fig. 4

Control DNA



DNA "triplet"



■ specific sequence for DR1 and 2
* pseudogene

INTERNATIONAL SEARCH REPORT

International Appl. No. PCT/US91/07308

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ⁶

According to International Patent Classification (IPC) or to both National Classification and IPC
 IPC(5): C12Q 1/68; C07H 15/12; C12N 15/00
 U.S. Cl.: 435/6; 536/27; 935/77, 78

II. FIELDS SEARCHED

Minimum Documentation Searched ⁷

Classification System	Classification Symbols
U.S. Cl.	435/6; 536/27; 935/77, 78

Documentation Searched other than Minimum Documentation
 to the Extent that such Documents are Included in the Fields Searched ⁸

III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹

Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
X	Proceedings of the National Academy of Sciences, Vol. 85, issued October 1988, Gyllensten et al, "Generation of Single-Stranded DNA by the Polymerase chain Reaction and Its Application to Direct Sequencing of the HLA-DQA Locin", page 7652-7656, see entire document.	1-21
X	Science, Vol. 230, issued 20 December 1985, Saiki et al, "Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia", pages 1350-1354, See entire document.	1-21
X	Science, Vol. 239, issued 29 January 1988, Saiki et al, "Primer-Directed Enzymatic Amplification of DNA with a Thermastable DNA Polymerase", pages 487-500, see entire document.	1-21

- Special categories of cited documents: ¹⁰
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search

28 January 1992

Date of Mailing of this International Search Report

10 FEB 1992

International Searching Authority

ISA/US

Signature of Authorized Officer

Jerome Massie

tf